

**Comments on 'Guidance for Industry
Patient-Reported Outcome Measures:
Use in Medical Product Development to Support Labelling Claims'
DRAFT GUIDANCE February 2006**

To: The Division of Dockets Management (HFA-305),
Docket number: 2006D-0044
<http://www.fda.gov/dockets/ecomments>
Food and Drug Administration,
5630 Fishers Lane
rm 1061
Rockville
MD 20852
USA

From: Clare Bradley, PhD, C.Psychol, FRSM, FBPSS
Professor of Health Psychology
Director of Health Psychology Research
Royal Holloway, University of London
Egham, Surrey
TW20 0EX
UK
Email: c.bradley@rhul.ac.uk

I have a particular interest in the FDA's draft guidance on patient-reported outcome (PRO) measures as I specialise in the design, development and use of such measures and license them to pharmaceutical companies, research organisations, academics and clinicians for use in clinical trials, other research and routine clinical practice. My measures include:

- the Diabetes Treatment Satisfaction Questionnaire (DTSQ) in it's status (DTSQs) and change (DTSQc) forms ¹⁻⁶ and related measures for other conditions including the HIVTSQ, RTSQ, RetTSQ, GHerpTSQ, ThyTSQ ⁷⁻¹¹, and the newly designed DTSQ-Teen and DTSQ-Parent. The DTSQs and c are fully linguistically validated in more than 60 language versions
- the Well-being Questionnaire (e.g. W-BQ12)^{3 12-14 15 16} generic measure of well-being is psychometrically validated for a range of populations including those who have diabetes (type 1 and type 2) macular disease and growth hormone deficiency and fully linguistically validated in 25 language versions
- the ADDQoL measure of the impact of diabetes on quality of life^{4 17} with related measures for other conditions including RDQoL, RetDQoL, MacDQoL, HDQoL, A-RHDQoL, ThyDQoL, ADDQoL Teen ^{11 18-25} and recently designed ADDQoL Jnr (for 5-8 year olds) and ADDQoL Jnr+ (for 9-12 year olds). The ADDQoL, MacDQoL and RetDQoL are linguistically validated in 16-25 language versions.

I welcome the FDA guidance as a much needed source of information about the standards required in PRO design, linguistic validation, psychometric validation and use and recognise that the guidance may be very useful in encouraging good practice.

I comment on issues in the order in which they first appear in the guidance and thereafter identify omissions that I ask be considered for inclusion.

Terminology, in particular ‘health status’ and ‘quality of life’: I welcome the use of the term ‘patient-reported outcome (PRO) measures’ in place of the ubiquitous and, usually inaccurate, use of ‘quality of life measures’. Previously ‘quality of life’ was used as an umbrella term to cover a wide range of PRO measures including measures of health status which are actually measuring quality of health and not quality of life. This distinction is important and failure to recognise the difference has led to some highly misleading conclusions and misguided policies. I emphasised the importance of distinguishing between quality of life and health status in a commentary in the Lancet in 2001²⁶ where I gave the example of the influential UK Prospective Diabetes Study (UKPDS) as one which used health status measures but interpreted their findings as if they were measuring quality of life²⁷. The UKPDS authors concluded that intensified treatment for Type 2 diabetes had no impact on quality of life and recommended widespread use of intensified treatment. In fact their findings showed there was no impact of intensified treatment on patients’ perceptions of the quality of their health and their quality of life was not measured. This is a very different conclusion and a far less desirable one than the one reached erroneously by the UKPDS authors.

Line 31 of the guidance defines a PRO as ‘a measurement of any aspect of a patient’s health status that comes directly from the patient...’. It seems that the FDA is here misusing the term ‘health status’, in the way that previously the term quality of life was misused, as an umbrella term to encompass a variety of other outcomes, which lines 35 and 36 suggest include symptoms, activities of daily living and quality of life. Thus quality of life measures are here conceptualised as a subset of health status measures instead of health status being (wrongly) seen as one of a range of quality of life measures and still there is no recognition of the importance of distinguishing between these two key concepts, health status and quality of life. If the term ‘health status’ is upgraded in this way to take over from ‘quality of life’ as an umbrella term we will have as much, if not more, confusion over terminology and, worse still, patient reported outcomes will come to be seen as measures of health as viewed by the patient. There is a danger that clinical trials will be satisfied with measuring patients’ reports of symptoms and will fight shy of measuring what the FDA are describing as ‘extremely complex concepts such as quality of life’. The great advantage of the term ‘PRO’ is that it is a neutral term that covers all patient reported outcomes including their satisfaction with their treatment, their well-being, their quality of life and their symptoms and health without needing an interim term such as ‘health status’ to limit the definition. If the FDA really feels the need to describe the kind of outcomes that PROs refer to then they might consider ‘health and quality of life outcomes’ which is the phrase eventually agreed upon by the Bio Med Journal of which I am an editorial board member. This phrase makes it clear that health outcomes are one form of PRO and quality of life outcomes are another and both are important but different, an issue discussed in the first editorial of the journal written by myself and the editor, Marcello Tamburini²⁸. My only concern with this phrase, is that it is not clear to me where patient satisfaction with treatment fits in. Treatment satisfaction is not a measure of health status but nor is it, strictly speaking, a measure of quality of life even though it may affect quality of life. I am reassured somewhat by the fact that the phrase refers to ‘quality of life outcomes’ and not ‘quality of life’ per se or ‘quality of life measures’. However, in many ways it is clearer and simpler to refer just to ‘Patient Reported Outcomes’ and make it clear that these can include symptoms, health status, treatment satisfaction, well-being and quality of life measures.

Conceptualising quality of life: The guidance describes the concept of quality of life as ‘extremely complex’. If we try to define what quality of life is in a way that is appropriate

for everyone it is indeed a very complex and perhaps an impossible task. However, if we follow the advice of Dick Joyce and define quality of life in terms of what the individual thinks it is²⁹ and measure it using individualised measures (e.g. ^{30 17}), it becomes a manageable, measurable and useful concept. It is very important for patients that clinical trialists do not duck the issue of measuring the impact of new treatments on the quality of life of individual participants in the trials and measure only the quality of their health. It is the bigger issue of quality of life that is most important to patients and it makes a great deal of difference to them if new treatments impair their quality of life or improve their quality of life. Only the patients can tell us how a treatment affects their quality of life using individualised PRO measures designed for the purpose (e.g. ^{17 18 20 23 19}).

Suggestions and recommendations: On lines 57 and 58 it is said that 'The use of the word *should* in Agency guidance documents means that something is suggested or recommended but not required'. I appreciate the good sense of this though in many instances whether something is recommended or not may depend on the particular measure, patient population or circumstances in which it is used. I would be inclined to qualify the 'is suggested or recommended' with 'is generally (or usually) suggested or recommended' to allow for exceptions to the usual rule.

Separate measures of adverse consequences and effectiveness: Lines 153-157 say 'Some PRO measures (e.g. health-related quality of life instruments) attempt to measure both the effectiveness and the side effects of treatment. PRO instruments that are used in clinical trials to support effectiveness claims should measure the adverse consequences of treatment separately from the effectiveness of treatment.' No explanation is given for this recommendation and it is unclear what the particular concern is here. Undoubtedly it is important for researchers to know about advantages and disadvantages of a treatment and detailed information about both is valuable for improving treatments. However, in developing individualised measures of the impact of diabetes and other chronic conditions on quality of life it is appropriate for patients to make holistic judgements about the impact of their condition including the treatment and any complications on separate aspects of their lives. Better that they weigh up the pros and cons themselves than have researchers make assumptions on their behalf. Patients are quite capable of making such judgements of impact even though they may be influenced by their view of the effectiveness of treatment and problems created by the treatment including side effects. We can and do measure these views separately but if we only use separate measures of symptoms and biomedical outcomes of effectiveness then we cannot know how to combine them to obtain a measure of their impact on the individual's quality of life. I have found it possible and desirable to ask the patient to tell us about the impact of their condition and its treatment on their quality of life. Lines 153 to 156, without explanation, appear to contradict my research experience.

Quality of life in the taxonomy of PROs: Quality of life does not appear in Table 1. The use of the term 'Overall health status' rather confirms my concern that the FDA is substituting global misuse of the term 'quality of life' for global misuse of the term 'health status'. It is very important to recognise that health status measures are needed but they are not everything. Quality of life measures are an essential subset of PRO measures for which health status measures provide no substitute.

Modification of PRO instruments: I welcome efforts to discourage users of established validated instruments from tinkering with the wording of questionnaires unnecessarily while referring to the validation of the original instruments as evidence for the modified

instrument's validity and reliability (lines 176-181). However, with some instruments, such as the DTSQc, it is necessary to modify the instructions to relate specifically to the conditions of the clinical trial in which it is being used and we now have considerable evidence to show that the psychometric properties of the DTSQc remain robust to such changes^{6 5}. I encourage users of the DTSQc to check the psychometric properties on each new use but would not go as far as to say that each new use (with modified instructions) should be treated as if it is a new measure.

Use of single item measures of a general concept: (lines 212-225) While it may often be true that 'a single-item PRO instrument is usually unable to provide a complete understanding of the treatment's effect because a single item cannot capture all the domains of the general concept.' (lines 214-6), my experience of a single-item measure of quality of life (*In general my present quality of life is...excellent through to extremely bad*)^{11 18 19 20 21 23 25} suggests such an item can be valuable alongside other measures (eg ^{31 32}). Indeed a single item such as this may capture all domains of the general concept far better than an inadequate subset of items can do (e.g. without the single thermometer item, the brief health status tool such as the EQ5D³³ would be completely inadequate (instead of just inadequate) for measuring health status in the many people with Type 1 diabetes because its limited domains include some that are not impacted at all for most of this population (e.g. mobility, self care, usual activities, and pain/discomfort) while issues of major importance, like the problems of nocturnal hypoglycaemia, and of dietary restriction are omitted.)

Responses to an individual item within a questionnaire are not to be allowed to support a claim for improvement in the individual item domain unless further items were developed to measure the concept. I can understand this if there was no significant improvement in responses to the individual item, even though there was significant improvement to the overall score on the questionnaire, to which the individual item contributed. However, if the individual item also improved significantly it would seem appropriate that such an improvement be used to support a claim. It is often the case that the treatment satisfaction score obtained by summing the 6 treatment satisfaction items on the DTSQ shows significant improvements with a new treatment and that one or more of the items that make up the measure also improve significantly. If the item concerning the convenience of treatment is the only one to improve significantly it is not unreasonable to claim that improvements in treatment satisfaction were mainly due to the convenience of the treatment without the need to spend years developing a separate measure of convenience of treatment with multiple items measuring different aspects of convenience.

Hypothetical responses to items: Lines 302-3 say that 'items that ask patients to respond hypothetically ...are not recommended'. The examples given make good sense and I would support the FDA's specific recommendations. However, there are circumstances where we have found that items that elicit hypothetical responses are appropriate and valuable. In the ADDQoL individualised measure of the impact of diabetes on quality of life, people with diabetes are asked to respond to items such as 'If I didn't have diabetes my working life would be...' with response options ranging from 'very much better' to worse. In initial qualitative work it was found that such items elicited more responses indicative of negative impact of diabetes on quality of life than did items without the hypothetical element such as 'Because of my diabetes, my social life is...'. The latter, more direct approach appeared to trigger optimistic coping responses fostered by messages emphasised by support organisations such as Diabetes UK and

the American Diabetes Association which encourage people with diabetes to feel that most if not all their desires and ambitions can be fulfilled even though they have diabetes. Thus people would be likely to reply to the straightforward item asking about the impact of their diabetes on their social life that their diabetes has no impact on their social life. However, when we followed up such questions by asking 'But if you didn't have diabetes how would that affect your social life?' people would reply 'Ah well, yes it would be easier, I wouldn't have to worry about how long we were going to be in the bar before we went on to eat dinner and whether I could risk injecting insulin 30 minutes before I ate or if the food might be late in coming and I would become hypoglycaemic and upset my friends by becoming irritable'. Thus we have found that people who have spent their lives being as positive as they can be about their chronic condition can usefully be encouraged to give more realistic responses stripped of the usual positive spin when we use the hypothetical question. It is important to note that where the aspect of life under consideration is not applicable to everyone (e.g. working life or sex life) there is the opportunity to skip the item. I would be glad if the FDA would ensure that their appropriate caution about some hypothetical items does not result in a blanket boycott of questionnaires which include hypothetical items.

Comparison of present state with an earlier state: In lines 339 to 343 the FDA warns against instruments that rely on patients' memory in recalling experiences over a period of time: 'It is usually better to construct items that ask patients to describe their current state than to ask them to compare their current state with an earlier period..'. While this may often be good advice there are exceptions. Where measurement of patient satisfaction with treatment is concerned we often find that patients report being very satisfied with their current treatment until they experience a better treatment and then they want to be able to say that they are much more satisfied with the new treatment. If they have been given a status measure of treatment satisfaction at baseline and given optimum responses showing they are very satisfied, they will not be able to respond any more positively at follow up when they are using a new treatment with which they are much more satisfied. It was to overcome such ceiling effects with the DTSQs status measure of satisfaction with diabetes treatment that I designed and developed the DTSQc measure of change in treatment satisfaction for use at follow up. This allows patients to say that they were very satisfied at baseline with the treatment they were using prior to the trial but are very much more satisfied with the new treatment they experienced within the trial. The DTSQc is also valuable in cross over trials at the end of each treatment period. I have encouraged users of the DTSQc to design their trials so that the period over which respondents are asked to think back is a straightforward time period such as 12 months or 6 months and not overly precise periods such as 23 weeks. We also tailor the instructions to remind patients of the particular treatment change (e.g. 'when you started to use inhaled insulin) when trial designs involve complexities like screening periods. We are finding that the DTSQc provides valuable data when used in addition to the DTSQs that overcomes ceiling effects that are sometimes found when the DTSQs is used alone^{6 5}.

Asymmetric response options: Lines 367-369. It is suggested quite appropriately that response options should not bias the direction of responses. However, the example given suggests that by offering one negative choice, one neutral choice and two or more positive choices on a scale, it will make it more likely that patients will respond that they feel or function better. I have actually changed the symmetric response options originally used in the ADDQoL to asymmetric response options because respondents rarely used the response options which indicated that their quality of life would be worse if they didn't

have diabetes as few people see any benefits of having diabetes. Instead of having seven response options, *very much better*, *much better*, *a little better*, *the same*, *a little worse*, *much worse* and *very much worse*, the three response options indicating that life would be worse without diabetes are now reduced to a single response option, *worse*. Thus we now have a total of 5 response options: *very much better*, *much better*, *a little better*, *the same*, *worse*. This works well. We have to have one worse option as a small minority of people do want to indicate benefits of having diabetes (e.g. they are physically fitter because they do more exercise than they would if they didn't have diabetes; their social life has improved through membership of Diabetes UK) but we don't need more than one option. The change to asymmetric response options was first made to the MacDQoL and RetDQoL for people with macular disease and diabetic retinopathy because it was particularly important that people with visual impairments should not be given more to read than absolutely necessary. Subsequently the change was also made to the ADDQoL in the interests of simplifying the questionnaire. In the case of the ADDQoL, MacDQoL and RetDQoL the numbers using the *worse* response options were small anyway and we haven't systematically examined the change before and after reducing the number of response options but we can do this. It is possible that the number of people using the worse options decreased slightly but if so it is more likely because they had previously used the *worse* end of the scale in error when they meant to use the *better* end when there were equal numbers of response options at either side of the scale. We may therefore see a small improvement in internal consistency with the asymmetric response options alongside any decrease in use of the *worse* end of the scale.

Development of format, instructions and training: Lines 390-394 specify changes to an instrument that might alter or influence patients' responses. I fully support the statement (lines 396-398) that 'It is important that the PRO instrument format used in the clinical trial be consistent with the format that is used in the instrument validation process. *Format* refers to the exact appearance of the instrument...'. The setting of a standard questionnaire into a commercial organisation's format with e.g logo, questions situated within a frame, additional material in footer or header can frustrate some respondents e.g. visually impaired people, and distract others. We have been careful to format our vision-related questionnaires (e.g. MacDQoL and RetDQoL) to optimise ease of completion. When they are adapted to accommodate a company's questionnaire 'livery' the result can be a questionnaire that may be less easy for a visually impaired person to complete, leading to missing data or even withdrawal from the study unless the problem can be overcome by the use of interviewers. Companies can be very insistent that their format be adhered to but it may be to the detriment of the project. There is now pressure to produce versions of my questionnaires in a revised format specified by one of the major pharmaceutical companies to facilitate electronic regulatory submissions to the FDA. It is not clear to me how much the revised format is stipulated by the FDA and how much is driven by the preferences of the pharmaceutical company. The changes required will probably not cause problems in completion. However, the changes are fairly extensive and psychometric analysis will be needed to establish equivalence to the original version developed. I don't think it is necessary to treat the new version as if it is a new questionnaire but it will be necessary to conduct confirmatory factor analysis to check that the psychometric properties are not affected by the changes made.

Response burden Line 449 stipulates 'Questions that patients are unwilling to answer' may contribute to respondent burden. It is possible that patients are coerced into

responding to questions they do not find acceptable by the setting in which they are asked. In research with visually impaired patients, time trade-off (TTO) questions are sometimes asked during hospital appointments after the patient has had his/her eyes dilated prior to the consultation with the ophthalmologist. Patients may feel vulnerable and disempowered in this situation and unable to decline to respond to difficult questions to which they may have personal or moral objections. Response rates may therefore erroneously suggest that patients were freely willing to participate. Mitchell and Bradley³⁴ reported a poor response rate for TTO questions when they were administered by telephone to participants who were in their own homes at a time convenient to the participants and comments from some of those people suggested that they found the questions ridiculous, too hypothetical or objectionable for religious or moral reasons³⁴. These same participants completed the MacDQoL measure without any objections or difficulties³⁵ showing that problems were specific to the TTO questions.

Minimum important difference (MID): Table 4. I agree that it can be helpful to consider the MID for clinical measures which are intermediate outcomes that may not be important outcomes for the patient in themselves but only in so far as they are predictors of other outcomes that are important (e.g. HbA1c measures of blood glucose control in diabetes). With some PRO measures that ask about symptoms, health status or visual functioning without asking about the importance of the issue in question for the patient, it may also be important to determine MID. However, when PROs have been designed to measure the importance of the aspect of life in question for the individual's quality of life as well as the impact of the condition on that aspect of life, there is no need for any additional information about MID. A statistically significant difference between treatment groups on the ADDQoL, RetDQoL, MacDQoL and related measures shows that the differences are not due to chance and we know from the patient's own responses how important the issues are to them. With treatment satisfaction items it is always possible that items have been selected to ask about aspects of treatment that are not considered important to the patients and it is necessary to establish how the items were generated and whether there are data on the importance of these items to patients. In the case of the –TSQ measures modelled on the DTSQ for diabetes, items were originally designed on the basis of qualitative work to establish the characteristics of the treatment important for patient satisfaction and dissatisfaction. In addition at least one study has been conducted to determine empirically the importance of each item in the DTSQ for the patient responders and all items were shown to be of high importance (Singh PhD thesis, Royal Holloway, University of London, in preparation). In such circumstances where the importance of the items is well established in design work and subsequent studies and the overall outcome being measured is not an intermediate outcome but one that is important to patients in its own right, it may be enough to show that there is a statistically significant difference between the treatment arms on this outcome measure. Of course a significant difference in the outcome will be easier to obtain if the sample size is great and it will be helpful to know the size of the effect in interpreting the results and compare this across studies. In sum, a statistically significant difference on measures of treatment satisfaction and well-being that have been designed to measure issues of importance to patients will necessarily be an important difference. So too will be a statistically significant difference on an individualised measure of the impact of a condition on quality of life, where the importance of an aspect of life for an individual's quality of life is part of the assessment.

I was not impressed with the list of ways in which people have attempted to derive MIDs that the FDA has reviewed and the comments made by the FDA suggested that they

have serious reservations too (lines 554-564). I also have major concerns about the first method outlined (551-554) which was not commented on by the FDA. This method involved mapping changes in PRO scores to clinically relevant and important changes in non-PRO measures and suggests that PRO measures be judged by their similarity to non-PRO measures such as spirometry scores in asthma. If this were appropriate there would be no purpose in measuring PROs at all. We might as well rely entirely on objective clinical measures. However, while it may be appropriate to expect some PRO measures such as those measuring health status or visual function to map onto clinical measurements, it is not appropriate for other PROs such as patient satisfaction or well being or the impact of the condition on quality of life which depend on much more than the clinical outcomes achieved. These latter PROs will depend on the demands of treatment and the extent to which the treatment can be adapted to suit the individual without damage to quality of life. It is important that we should be able to measure these PROs without being required to show that they map onto non-PROs! Indeed, it is perfectly possible that despite bringing about improvements in clinical outcomes a new treatment causes greater negative impact on treatment satisfaction and quality of life and, if so, patients are unlikely to be able to maintain clinical improvements in the long term.

The assessment of measurement properties This section includes many important issues but one that is notably absent is any requirement for empirical evidence for the structure of a measurement that is deemed to have subscales. It is a common practice to design subscales to measure concepts that are theoretically rather than empirically derived and then to establish the measurement properties of the subscales without seeking evidence from principal components analysis or other such methods that the instrument does indeed split into the subscales as expected. The authors of the chapter on the DQOL (Diabetes Quality of Life) measure did not provide evidence for the subscales which supposedly made up the measure although they provided suitable evidence for the measurement properties for each of the subscales³⁵. Subsequent work that has looked for supporting evidence for the presence of the intended subscales using factor analysis has been unsuccessful in finding such support³⁶. I think that it is important to encourage questionnaire developers to provide evidence for the structure of the questionnaire in terms of any identifiable subscales.

Modification of an existing instrument: I would welcome a stronger message to encourage pharmaceutical companies to use PRO measures exactly as supplied by scale developers and not to reformat them in their house style giving priority to logos and headers at the expense of font size and clarity. Whenever clients want to retype questionnaires it takes many rounds of review and requests for changes before all errors are removed and potential problems resolved with the original format preserved. I have never yet seen one of my questionnaires retyped without problems being introduced and resolving these problems is time consuming and difficult. On one occasion a major pharmaceutical company expected to be able to reformat a questionnaire in their house style despite having signed a licence agreement requiring that no changes be made to the questionnaire. They did not think of the style modifications as 'changes to the questionnaire'. See also related points in paragraphs above 'Modifications of PRO instruments' and 'Development of Format, Instructions and Training'.

Protocol: Line 711 stipulates that the protocol include the 'exact format and version of the specific PRO instrument to be administered'. While agreeing that this is essential for judging the suitability of the PRO I also find it essential to include a 'For Information

Only' banner across versions of my PRO measures to be included in protocols and ethics committee submissions to reduce the risk of unauthorised use. It would be helpful if the FDA would explicitly acknowledge that such banners are acceptable providing the whole of the questionnaire can be seen through it.

Linguistic validation (LV) of PRO measures: I have already noticed that some major pharmaceutical companies who have previously paid only lip service to the need for linguistic validation, are now accepting that this is a task for specialists that will take 5 months and may cost as much as 20,000 Euros for one questionnaire to be fully linguistically validated in one language by specialists in the field. I believe that awareness that the FDA was preparing guidance has encouraged awareness of the need for professional LV work. I have long collaborated with Mapi in Lyon on such work and work closely with them to help ensure excellent results. Some pharmaceutical companies want to use very much cheaper competitors who promise to complete the work in a fraction of the time Mapi take, but experience has shown that Mapi's high standards cannot be met at such speed. I think it would be helpful to provide rather more guidance on the quality of LV work required to produce good translations of PRO measures. In particular it would be helpful to note that it is good practice for the developer of the measure to be closely involved in the LV work. I employ a full-time linguist to manage my collection of translations and she and I are actively involved in LV work on my questionnaires.

Blinding and randomisation: I must take issue with the statement that 'open-label studies, where patients and investigators are aware of assigned therapy, are rarely credible.' (line 717-8). In chronic disorders such as diabetes, all participants in trials will receive active treatment and the issue is more often whether they receive a new treatment or continue with an existing treatment rather than whether they receive active treatment or placebo. New treatments may carry risks and possible unwanted effects as well as benefits and it is not appropriate to assume that patients will always be more positive about a new treatment than about an old treatment. It is said on line 721 that 'Every effort should be made to assure that patients are masked to treatment assignment throughout the trial'. In practice this may mean that patients are asked to use two treatments, one of which is a placebo. This places additional demands on the patients that do not reflect the clinical realities of either treatment and render the trial unsuitable for evaluating the impact of treatments on patient satisfaction or quality of life. While I agree that 'The impact of unblinding is important to consider in the interpretation of study results' (line 723) it is equally important to consider the impact of blinding on study results. Blinding should not be assumed to be universally desirable and in itself can distort study results.

Line 726 suggests that 'questions that ask how patients' current status compares to baseline seem likely to be more influenced by unblinding (optimism can readily be expressed as a favourable comparison) than questions about current status (which requires a current assessment, not a statement about duration)' (Do you mean differences rather than duration here?) It is particularly frustrating that there is no reference given for evidence for this point. In my experience of using the DTSQs (status measure) and the DTSQc (change measure) we often see that the DTSQc shows greater improvements in satisfaction with treatment than are shown by the DTSQs. However, separate analysis of patients who scored at or near ceiling on the DTSQs at baseline and patients who had more room to show improvement in

satisfaction showed clearly that ceiling effects were limiting the benefits shown when the status measure alone was used and the DTSQc provided a more accurate representation of the benefits patients experienced⁵. It is possible that other studies showing fewer benefits with status measures than with change measures are in fact underestimating the benefits of treatment due to ceiling effects with the status measures that are overcome by using change measures.

Avoiding missing data due to withdrawals: It can be valuable to include an early interim PRO data collection point in a clinical trial, say at 6 or 8 weeks, after patients have overcome any initial teething problems with their treatment, so that there is more likely to be an on-treatment endpoint measure for participants who drop out before the planned end of the study. Such data, together with subsequent interim data points, are anyway useful to track changes in such outcomes as treatment satisfaction during the course of a trial. A steady increase in treatment satisfaction across a 6 month or one year trial (e.g.³⁷) removes any concerns that treatment satisfaction may have increased initially on a new treatment as a result of optimism that subsequently turns to disappointment with a consequent reduction in treatment satisfaction. A progressive increase in patient satisfaction provides reassurance that responses are based on experience rather than expectations.

Statistical considerations for patient-level missing data: Line 1004 refers to imperfect strategies that 'try to predict missing outcomes for a patient who has withdrawn from the trial using data from subjects* who stayed in the trial and for whom all data have been collected'. Participants who withdraw from trials are likely to have worse scores on PROs such as treatment satisfaction measures than are those who continue in a trial and to impute missing values for those who withdraw from those who remain is likely to overestimate patient satisfaction. It would be much more informative to give the PRO measure to participants who withdraw early or to include interim data collections of PROs for use in endpoint analyses as suggested above.

***Subjects:** The British Psychological Society advises that the term 'subjects' not be used as it can cause offence and suggests to some potential participants that they may be subjected to unpleasant experiences. I now avoid the term and use 'participants', 'respondents' or 'individuals'. The FDA might wish to follow suit and avoid all use of the term 'subjects' in these guidelines.

Glossary

Health-related quality of life (HRQL) is defined as 'A multidomain concept that represents the patient's overall perception of the impact of an illness and its treatment. An HRQL measure captures, at a minimum, physical, psychological (including emotional and cognitive), and social functioning. Claiming a statistical and meaningful improvement in HRQL implies: (1) that the instrument measures all HRQL domains that are important to interpreting change in how the study population feels or functions as a result of treatment; and (2) that improvement was demonstrated in all of the important domains.' This definition would seem to allow for some health status measures (which measure quality of health) to be classed as HRQL measures (e.g. SF-36 for some patient groups) as well as condition-specific quality of life measures (e.g. ADDQoL^{4 17}). I think this encourages health status measures to be mislabelled as if they were quality of life measures (or health-related quality of life measures) when they are measures of the quality of health and creates problems of interpretation discussed above and elsewhere²⁶.

Point (1) above will exclude many generic tools which do not adequately assess the impact of specific conditions on aspects of life important for quality of life: this may be an important step forward. For example, the aspect of life measured by the ADDQoL that is most impaired by diabetes is *freedom to eat as I wish*, is not measured by any other quality of life measure that I am aware of. Awareness of this major influence of dietary restrictions on quality of life led to the evaluation of the DAFNE (Dose Adjustment For Normal Eating) approach to insulin treatment for diabetes with major benefits to quality of life, treatment satisfaction and glycaemic control³². The DAFNE approach was supported by the recent National Service Framework for Diabetes in the UK and the Department of Health has funded roll out of the approach nationwide.

Point (2) above: I would take issue with the suggestion that improvement needs to be demonstrated in all of the important domains in a HRQL instrument. First because improvement can only be demonstrated in domains where deficits are apparent to start with, however important the domain may be, and we cannot expect that deficits will always be found for all important domains in all uses of a questionnaire. Secondly it seems unreasonable to expect to see benefits for all important domains even if there were deficits to start with. The outstandingly successful DAFNE approach did not achieve significant improvements for all the domains of the ADDQoL even though it showed significant benefits on the overall score and on many specific domains³². Just as we wouldn't reject a treatment because not all blood measures showed improvements so too we should not reject a treatment because not all domains of a PRO measure improve.

Quality of Life is defined in the glossary as 'A general concept that implies an evaluation of the impact of all aspects of life on general well-being. Because this term implies the evaluation of nonhealth-related aspects of life, it is too broad to be considered appropriate for a medical product claim'. Surely it should be the ultimate aim of a treatment to benefit patients' quality of life? We may sometimes have to settle for reducing the damage done to quality of life by a medical condition such as diabetes but we will still need a definition of quality of life. The ADDQoL provides an overview question about quality of life *per se* as well as measuring the impact of diabetes on quality of life and there have been at least two studies which have shown significant benefits to quality of life on the overview present quality of life item as well as reductions in the negative impact of diabetes on quality of life^{31 32}. The definition of quality of life that I use and have used as a basis for the design of my ADDQoL and related measures is that quality of life is what the individual thinks it is. The individual's view of their quality of life may indeed include aspects of life that are not health related although as medical conditions become more severe and/or their treatment becomes more demanding and/or invasive, the aspects of life that are not health related diminish. I believe it is essential that we face up to the importance of defining and measuring quality of life *per se* and do not duck this issue or confuse matters further by defining quality of life in terms of well-being which is a term that is not defined in the glossary! There are many measures of well-being which typically include subscales to measure depression and anxiety, energy, and, sometimes, positive well-being (e.g.^{12 16}). When a person is depressed and anxious their quality of life is also likely to be impaired. However, someone who is not depressed or anxious may nevertheless feel that their quality of life would be much improved if they didn't have diabetes. Thus measures of well-being are no substitute for measures of quality of life. I recommend that the FDA adopt a simple patient-centred definition for the concept of quality of life – quality of life is what the individual concerned thinks it is²⁹ and encourages the considerable efforts made to date to measure individualised quality of life (e.g. the SEIQoL³⁰) and the impact of medical conditions on individual's quality of

life (e.g. the ADDQoL¹⁷) which have already been welcomed by several reviewers despite the first publication of the ADDQoL being only six years old.^{38 39 40}

Copyright: No mention is given of the need to respect copyright in questionnaires. Even major pharmaceutical companies have sometimes translated and used questionnaires without a licence and there needs to be clear guidance about this issue. Some authors of questionnaires have unwittingly caused confusion by declaring their measures 'in the public domain' and not coordinating linguistic validation work or improvements such that we find a proliferation of slightly different versions. If authors do not want to accept the responsibility for managing the questionnaire, they might be encouraged to subcontract that task to another organisation that specialises in such work (e.g. Mapi Research Institute in Lyons). There has recently been an increased demand for linguistic validation certificates documenting the procedures used in the linguistic validation work to produce new language versions of questionnaires and I believe this has been driven by the demands of the regulatory bodies. It certainly seems to be a useful way of discouraging unauthorised translations and ensuring that only authorised translations are sought and used. The FDA is in an excellent position to encourage good practice in obtaining PROs from copyright holders and discouraging unlicensed use.

References are needed: Lack of references to the examples given in the FDA guidelines makes it very difficult to accept a number of recommendations which appear to be over generalising from specific cases when that is not justified. For example, line 726 says 'questions that ask how patients' current status compares to baseline seem likely to be more influenced by unblinding (optimism can readily be expressed as a favourable comparison) than questions that ask about current status..' This may be a misinterpretation of ceiling effects leading to smaller changes with status measures than are seen with change measures. My investigations of such measures lead me to believe that the change measures give the more accurate reflection of patients' experiences as discussed above in the section on Blinding and Randomisation. A colleague who attended a meeting to discuss these guidelines with representatives from the FDA commented that when questions were raised about generalities made in the guidelines the answers always started with 'It depends...'. It would seem to make very good sense to acknowledge that many of the general points will depend on the particular circumstances and the provision of references would recognise the importance of providing evidence for recommendations and leave open the possibility that there may be other evidence which would support different conclusions.

Access to questionnaires, questionnaire guidelines and references: I will be glad to provide e-copies of any of my questionnaires listed on page 1 of these guidelines, and associated scoring instructions, user guidelines, and reference lists and can be contacted at the address in the footer or by email at c.bradley@rhul.ac.uk.

Summary: I believe that the FDA guidelines have already had an impact in encouraging good practice in the use of PROs. There are, however, important improvements that need to be made to the guidelines, particularly in the use of health status and quality of life terminology. It is essential to distinguish between health status and quality of life and to use both terms. Nothing is to be gained and a great deal will be lost if the term quality of life (which has been misused as an umbrella term in the past) is abandoned and replaced with the term health status. Patients want us to consider their quality of life as well as their health. To abandon the term would be to forget about their quality of life and focus only on their health. Patients are well able to tell us what quality of life means to

them and to rate the impact of a condition on their quality of life if we use individualised quality of life measures (e.g.³⁰) and individualised condition-specific quality of life measures (e.g.⁴) to allow them to do so. Although my experience with PRO measures would support many of the recommendations in the guidelines there are others that I would not fully agree with or would contradict on the basis of my own research evidence. I have provided references to that research and hope that the FDA will feel able to do the same when they finalise these guidelines.

References

1. Bradley C, Lewis KS. Measures of psychological well-being and treatment satisfaction developed from the responses of people with tablet-treated diabetes. *Diabetic Medicine* 1990;7:445-451.
2. Bradley C. The Diabetes Treatment Satisfaction Questionnaire: (DTSQ). In: Bradley C, editor. *Handbook of Psychology and Diabetes: a guide to psychological measurement in diabetes research and practice*. Chur, Switzerland: Harwood Academic Publishers, 1994:111-132.
3. Bradley C, Gamsu DS. Guidelines for encouraging psychological well-being: report of a Working Group of the World Health Organization Regional Office for Europe and International Diabetes Federation European Region St Vincent Declaration Action Programme for Diabetes. *Diabet Med* 1994;11(5):510-6.
4. Bradley C, Speight J. Patient perceptions of diabetes and diabetes therapy: assessing quality of life. *Diabetes Metab Res Rev* 2002;18 Suppl 3:S64-9.
5. Bradley C, Plowright R, Stewart J, Witthaus E. Diabetes Treatment Satisfaction Questionnaire (change) in English and German evaluated in insulin glargine trials. *Diabetologia* 2000;43(1):A196.
6. Howorka K, Pumprla J, Schlusche C, Wagner-Nosiska D, Schabmann A, Bradley C. Dealing with ceiling baseline treatment satisfaction level in patients with diabetes under flexible, functional insulin treatment: Assessment of improvements in treatment satisfaction with a new insulin analogue. *Quality of Life Research* 2001;9:915-930.
7. Woodcock A, Bradley C. Validation of the HIV treatment satisfaction questionnaire (HIVTSQ). *Qual Life Res* 2001;10(6):517-31.
8. Barendse SM, Speight J, Bradley C. The Renal Treatment Satisfaction Questionnaire (RTSQ): a measure of satisfaction with treatment for chronic kidney failure. *Am J Kidney Dis* 2005;45(3):572-9.
9. Woodcock A, Plowright R, Kennedy-Martin T, Hirsch A, Ffytche T, Bradley C. Development of the new Retinopathy Treatment Satisfaction Questionnaire (RetTSQ). *Proceedings of Vision 2005; International Congress Series* 2005;1282:342-346.
10. Taback N, Bradley C. Validation of the Genital Herpes Treatment Satisfaction Questionnaire (GHerpTSQ) in Status and Change Versions. *Quality of Life Research* 2000;9(3):332.
11. McMillan CV, Bradley C, Woodcock A, Razvi S, Weaver JU. Design of new questionnaires to measure quality of life and treatment satisfaction in hypothyroidism. *Thyroid* 2004;14(11):916-25.
12. Bradley C. The 12-Item Well-Being Questionnaire: origins, current stage of development, and availability. *Diabetes Care* 2000;23(6):875.

13. Pouwer F, van der Ploeg HM, Ader HJ, Heine RJ, Snoek FJ. The 12-item well-being questionnaire. An evaluation of its validity and reliability in Dutch people with diabetes. *Diabetes Care* 1999;22(12):2004-10.
14. Pouwer F, Snoek FJ, Van der Ploeg HM, Ader HJ, Heine RJ. The Well-Being Questionnaire: evidence for a three factor structure with 12 items (W-BQ12). *Psychological Medicine* 2000;30:455-462.
15. Mitchell J, Bradley C. Psychometric evaluation of the 12-item Well-being Questionnaire for use with people with macular disease. *Qual Life Res* 2001;10(5):465-73.
16. McMillan C, Bradley C, Gibney J, Russell-Jones D, Sonksen P. Well-being in adult growth hormone deficiency. *Health and Quality of Life Outcomes* 2006;In press.
17. Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. *Qual Life Res* 1999;8(1-2):79-91.
18. Bradley C. Design of a renal-dependent individualised quality of life questionnaire. *Advances in Peritoneal Dialysis* 1997;13:116-120.
19. Woodcock A, Bradley C, Plowright R, ffytche T, Kennedy-Martin T, Hirsch A. The influence of diabetic retinopathy on quality of life: interviews to guide the design of a condition-specific, individualised questionnaire: the RetDQoL. *Patient Educ Couns* 2004;53(3):365-83.
20. Mitchell J, Bradley C. Design of an individualised measure of the impact of macular disease on quality of life (the MacDQoL). *Qual Life Res* 2004;13(6):1163-75.
21. Mitchell J, Wolffsohn JS, Woodcock A, Anderson SJ, McMillan CV, Ffytche T, et al. Psychometric evaluation of the MacDQoL individualised measure of the impact of macular degeneration on quality of life. *Health Qual Life Outcomes* 2005;3(1):25.
22. McMillan CV, Bradley C, Gibney J, Healy ML, Russell-Jones DL, Sonksen PH. Psychological effects of withdrawal of growth hormone therapy from adults with growth hormone deficiency. *Clin Endocrinol (Oxf)* 2003;59(4):467-75.
23. McMillan C, Bradley C, M G, Martin F, Sonksen P. Preliminary development of a new individualised questionnaire measuring quality of life in older men with age-related hormonal decline: the A-RHDQoL. *Health and Quality of Life Outcomes* 2003;1(51).
24. McMillan CV, Honeyford RJ, Datta J, Madge NJ, Bradley C. The development of a new measure of quality of life for young people with diabetes mellitus: the ADDQoL-Teen. *Health Qual Life Outcomes* 2004;2(1):61.
25. McMillan C, Bradley C, Gibney J, Russell-Jones DL, Sonksen PH. Preliminary development of the new individualised HDQoL questionnaire measuring quality of life in adult hypopituitarism. *Journal of Evaluation in Clinical Practice* 2006;In press.
26. Bradley C. Importance of differentiating health status from quality of life. *Lancet* 2001;357(9249):7-8.
27. U.K. Prospective Diabetes Study Group. Quality of Life in Type 2 Diabetic Patients Is Affected by Complications But Not by Intensive Policies to Improve Blood Glucose or Blood Pressure Control (UKPDS 37). *Diabetes Care* 1999;22(7):1125-1136.
28. Bradley C, Tamburini M. Not-only-a title. Editorial. *Health and Quality of Life Outcomes* 2003;1(1).
29. Joyce CRB. Requirements for the assessment of individual quality of life. In: McGee HM, Bradley C, editors. *Quality of Life Following Renal Failure: Psychosocial Challenges Accompanying High Technology Medicine*. Chur, Switzerland: Harwood Academic, 1994:43-54.

30. McGee HM, O'Boyle CA, Hickey A, O'Malley KM, Joyce CRB. Assessing the quality of life of the individual: The SEIQoL with a healthy and a gastroenterology unit population. *Psychological Medicine* 1991;21:749-759.
31. Johnston P, Stephens J, Witthaus E, Bradley C. Improved treatment satisfaction and quality of life with insulin glargine +lispro compared with NPH insulin + regular human insulin. *The Endocrine Society's 86th Annual Meeting Program and Abstract Book*; 2004:397.
32. DAFNE. Training in flexible, intensive insulin management to enable dietary freedom in people with type 1 diabetes: the dose adjustment for normal eating (DAFNE) randomised control trial. *BMJ* 2002;325:746 - 749.
33. EuroQoL Group. EuroQoL - a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
34. Mitchell J, Bradley C. Measuring quality of life in macular disease: What use are utilities? *International Congress Series* 2005;1282:654-658.
35. Jacobson AM, The Diabetes Control and Complications Trial Research Group. The Diabetes Quality of Life Measure. In: Bradley C, editor. *Handbook of Psychology and Diabetes: A Guide to Psychological Measurement in Diabetes Research and Practice*. Chur, Switzerland: Harwood Academic, 1994:65-88.
36. Eiser C, Flynn M, Green E, Havermans T, Kirby R, Sandeman D, et al. Quality of Life in Young Adults with Type 1 Diabetes in Relation to Demographic and Disease Variables. *Diabetic Medicine* 1992;9:375-378.
37. Witthaus E, Stewart J, Bradley C. Treatment satisfaction and psychological well-being with insulin glargine compared with NPH in patients with Type 1 diabetes. *Diabet Med* 2001;18(8):619-25.
38. Polonsky WH. Understanding and Assessing Diabetes-Specific Quality of Life. *Diabetes Spectrum* 2000;13(1):17-22.
39. Garratt AM, Schmidt L, Fitzpatrick R. Patient-assessed health outcome measures for diabetes: a structured review. *Diabet Med* 2002;19(1):1-11.
40. Greenhalgh J, Long AF, Brett AJ, Grant MJ. Reviewing and selecting outcome measures for use in routine practice. *J Eval Clin Pract* 1998;4(4):339-50.